

---

# Model Sensitivity to User Load

Robert Howard

---

21 February 1996

# Roadmap



## Overview

### Comparison of the Push loads versus the Pull loads on the system

- This is a follow-on to the previous Push vs Pull comparison
- Uses Blocking Buffer

### Sensitivity Studies

- Volume Composition (number of requests vs. size of request)
- Request type (number of retrieves vs. number of manipulates)
- Decimation (number of files per manipulate request vs. decimation ratio)

## Summary

# Overview



**Hardware capacity requirements are sensitive to user pull assumptions.**

- Sensitivities to the assumptions must be understood because they may have significant impact on system sizing.

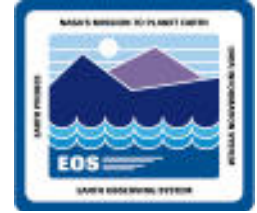
**User pull requirements need interpretation**

- Volume of data distributed is stated only as 2x the volume migrated/produced.

**This series of studies uses the dynamic model to assess and bound effects of the user pull assumptions**

- This presentation is a “in progress” look at these sensitivity studies
- These studies indicate the relative sensitivity of the system to the various parameters and assumptions. They do not necessarily imply an analytic formula that can be used in sizing.

# Requirements for supporting User Pull



**The August, 1995 Technical Baseline (attachment I-1) states:**

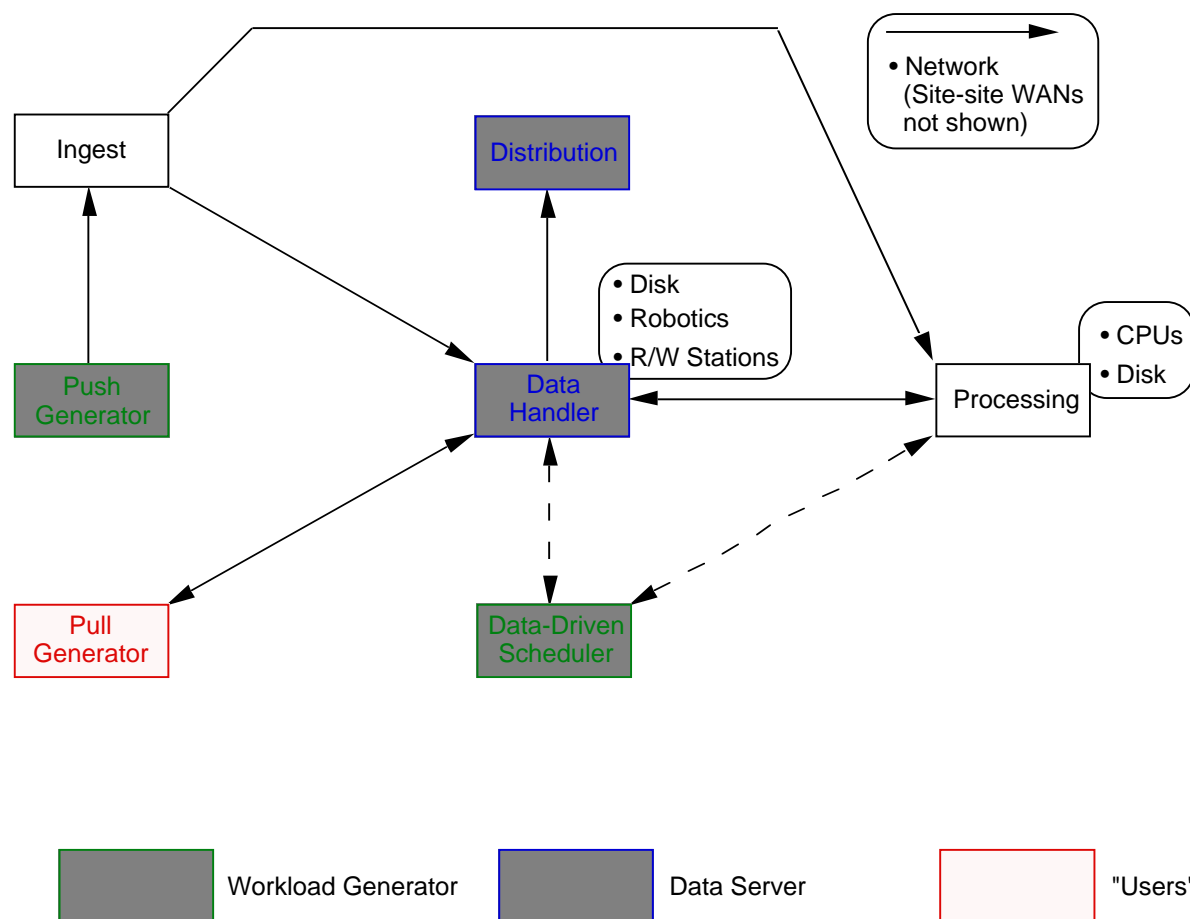
- The volume distributed is assumed to 2x the volume migrated/produced.
- DAS products and Product QA are included in the 2x distribution.

***Problem*—This doesn't address:**

- User pull per unit time
- Number of requests or files that compose the distribution volume
- Amount of subsetting and/or processing required to prepare the volume
- Ratio of volume extracted from archive to volume sent to user
  - referred to as the decimation ratio
- Length of time data is held on disk for user pull pickup

**Approach: Perform sensitivity analysis to assess effects of assumptions**

# Dynamic model context showing measurement points



- **Buffer used to store files before they are written to the archive**

**PGS Disk: Disk pool local to processing, in Gigabytes**

**R/W Stations:** Read/Write stations in the archive device

**avg: Long-term time Average**

**DSS: Data Server**

**DPS: Data Processing System (generation of standard products)**

## DMS: Data Management System

[illegible]

## August 1995 Baseline

## Blocking Buffer in use

## Reprocessing not in Push load

## DAO and MODIS Level 3 products not in Push load

## Resources from all DAACs summed together for presentation

### Quantity of processors normalized to 360 MFLOP processors

## Subsetting, etc. performed on separate string within DPS

**Disk utilization is in GB and Network utilization is in MB/second**



# Revisit of Push vs Pull

**Comparison is made in order to determine whether changes observed in the User Pull Sensitivity analyses have a significant effect on the total resources needed by ECS.**

**Earlier study (Archive Sensitivity) suggests that “large” and “small” DAACs should be separated for analysis.**

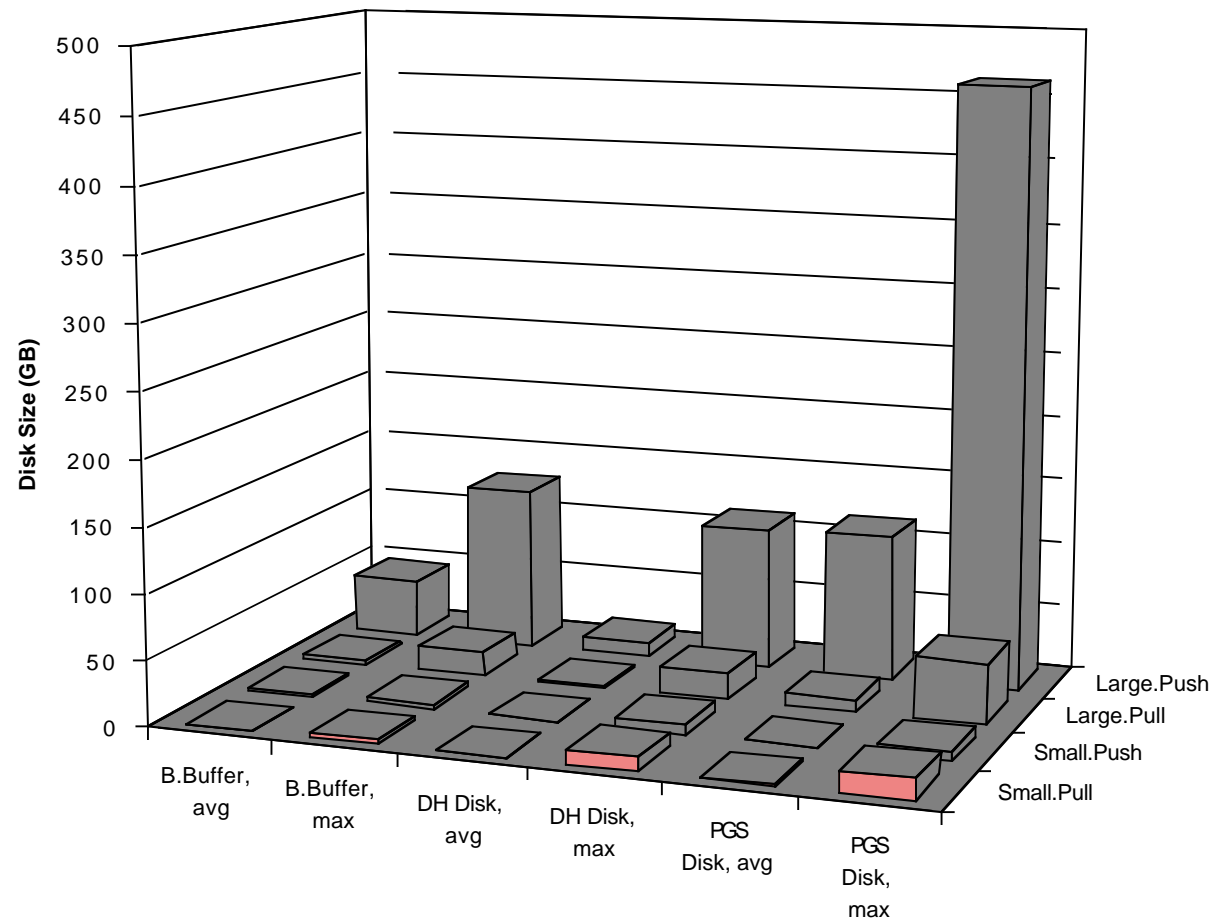
- **Large: EDC, GSFC, LaRC**
- **Small: ASF, JPL, MSFC, NSIDC, ORNL**
- **For analysis, the large sites were averaged and the small sites were averaged**

**Data presented for Push Only is first time processing only.**

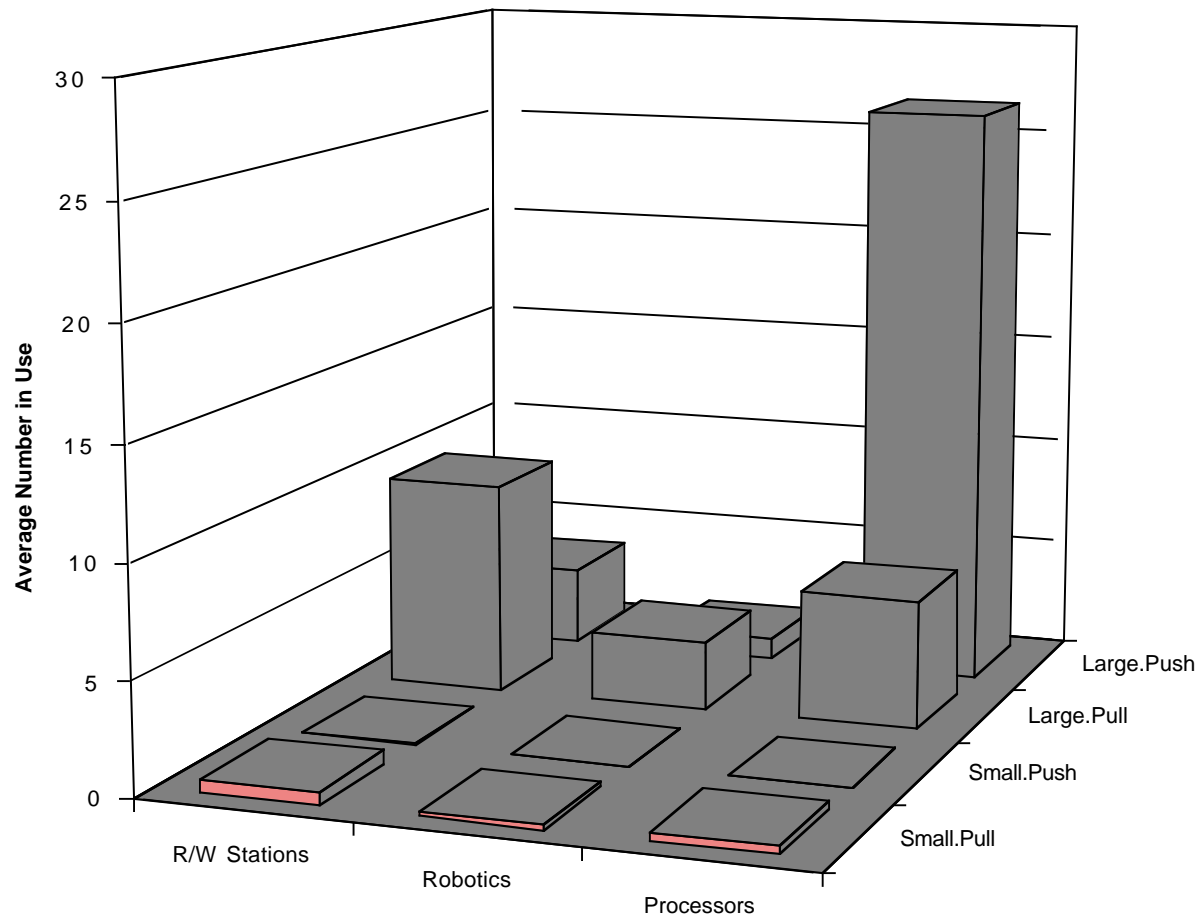
- **Multiplying “Push Only” requirements shown by phasing factor approximates capacity needed “on the floor” if “head-of-chain” reprocessing is used.**
- **Capacity phasing factor in 3rd Qtr 1999 is 2.2**



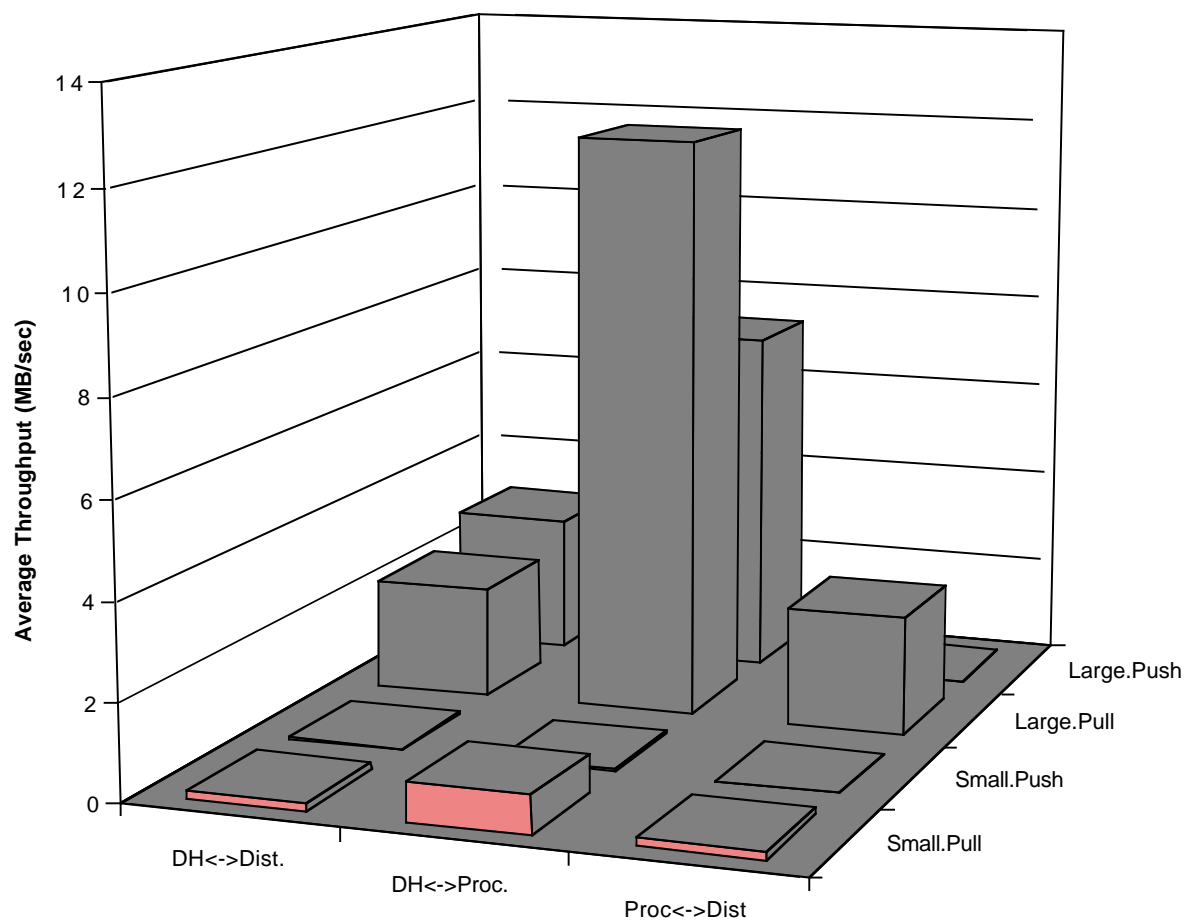
# Push vs Pull disk requirements



# Push vs Pull on a subset of resources



# Push vs Pull network requirements





# Push versus Pull Results

## Requirements for Working Storage (both DSS and DPS)

- At small sites, push and pull are in rough balance
- At large sites, working storage was dominated by push

## Requirements for Processors

- At small sites, push and pull are in rough balance
- At large sites, number of processors was dominated by push

## Requirements for Archive devices dominated by Pull

- Includes both Robotics and Read/Write Stations
- After phasing factor applied, Push is closer to balancing
  - Factor of ~3 needed to balance
- Reprocessing paradigm other than “head of chain” causes Push to dominate
- Previous result: removal of blocking buffer causes Push to dominate



# Push versus Pull Results (cont.)

## Network requirements mixed

- At “small” DAACs, Pull predominates
- At “large” DAACs, DSS $\leftrightarrow$ Dist’n is balanced between Push and Pull
- At “large” DAACs, DSS $\leftrightarrow$ DPS is dominated by Pull
  - After phasing factor applied, Push higher
  - If servicing manipulate requests is moved to DSS, Push will dominate
- DPS $\leftrightarrow$ Dist’n effected only by Pull (due to subsetting, etc.)
  - If servicing manipulate requests is moved to DSS, this logical link goes away

# Volume Composition Experiment



**Experiment compares archive operation based on differing number of requests and differing amount of volume distributed**

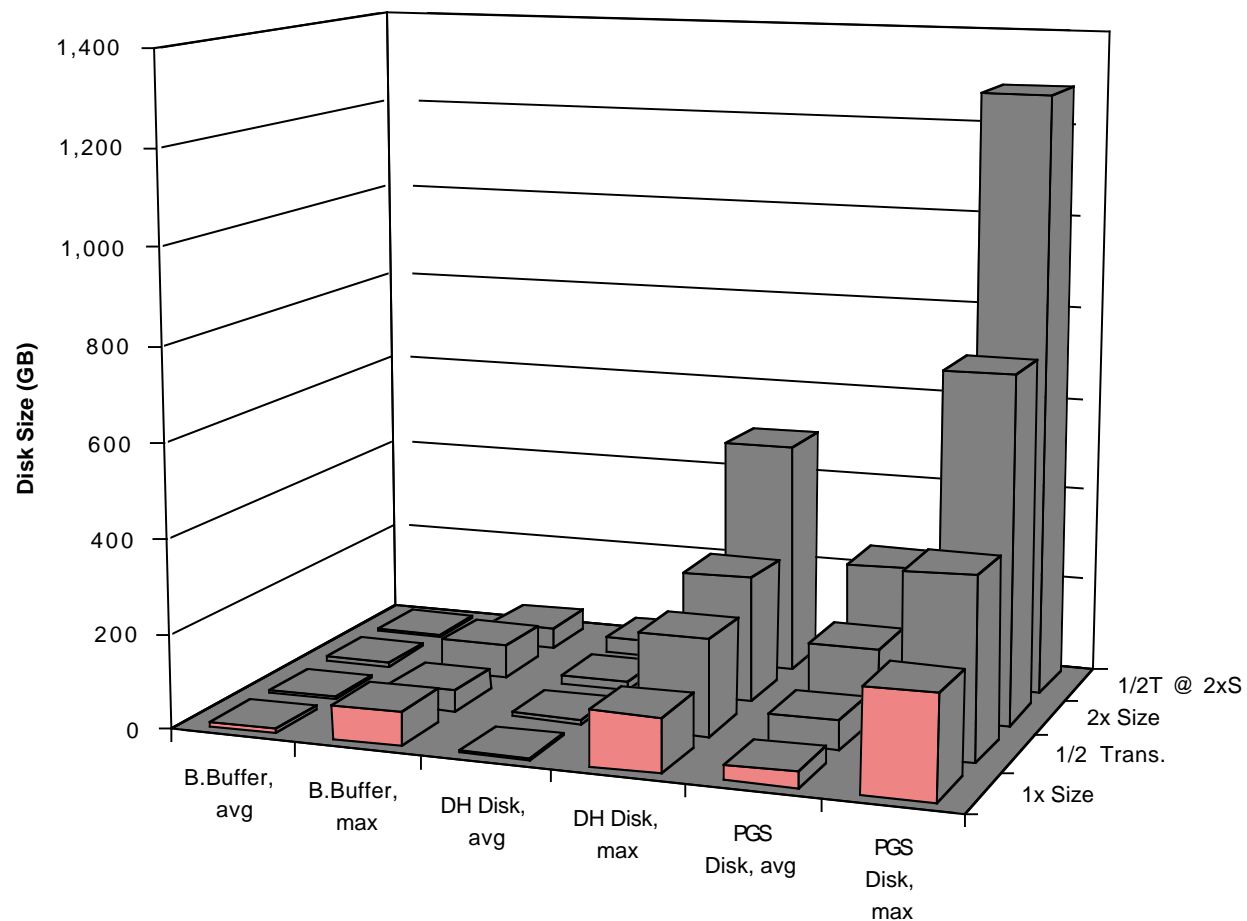
**Experiment set up as a 2<sup>2</sup> full factorial statistically designed experiment**

- **Baseline number of requests per unit time versus 1/2 that number**
- **1x file size versus 2x file size**

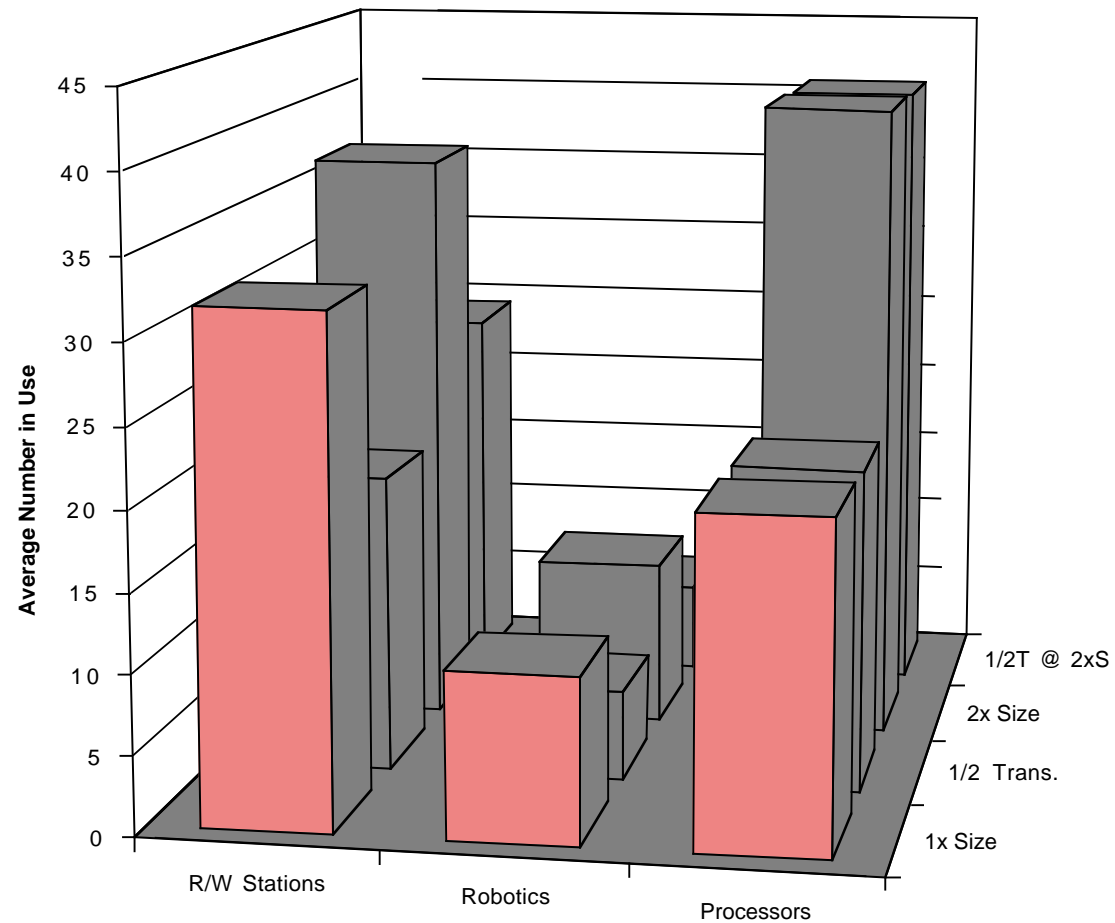
**While the experiment halved the number of transactions, they could have just as easily been doubled**

- **So far as the experiment is concerned the two situations are equivalent (i.e., underlying trends will be the same)**

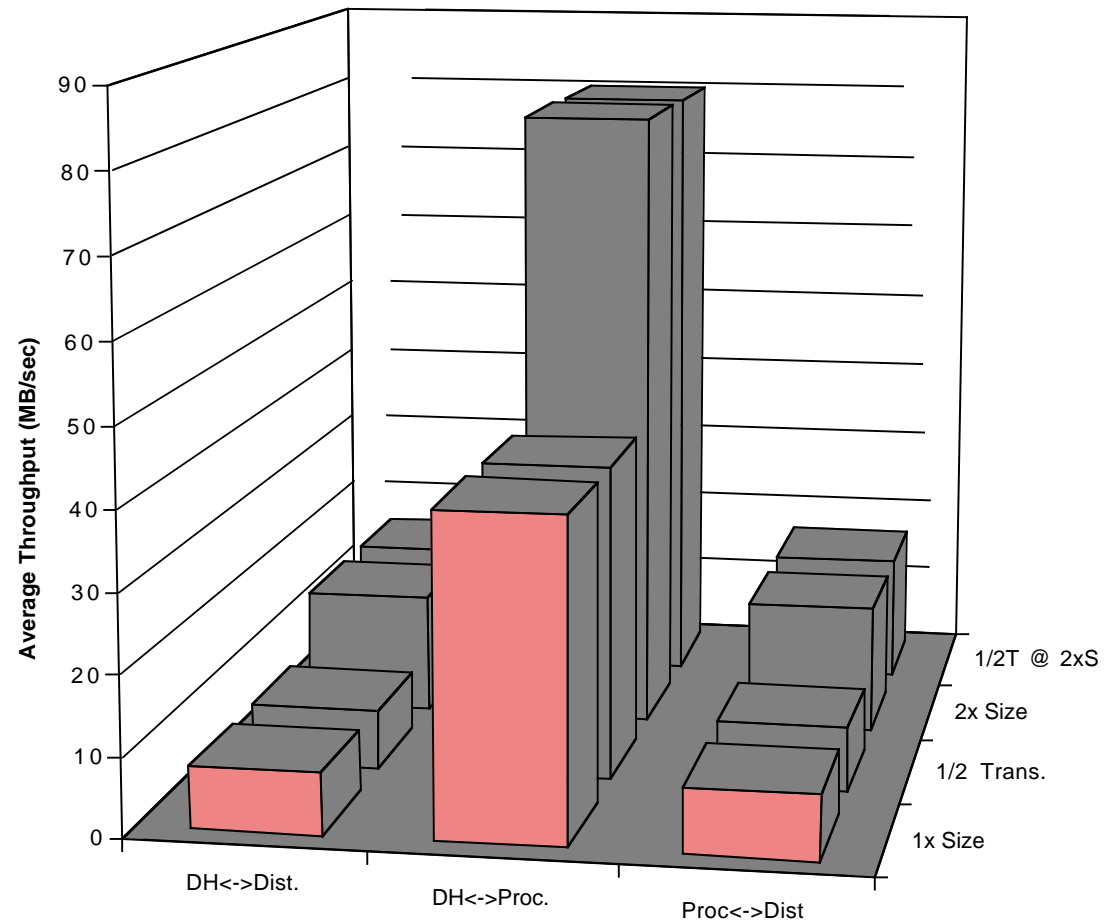
# Volume Composition disk requirement results



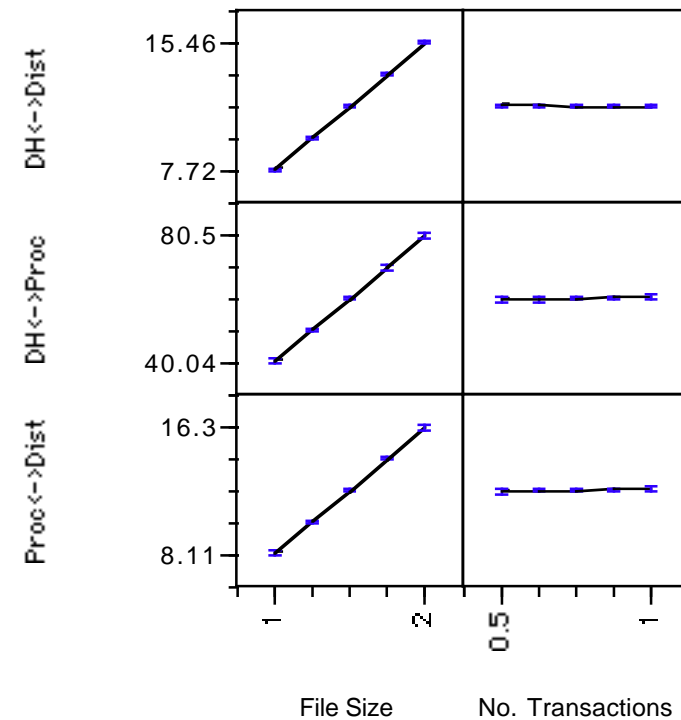
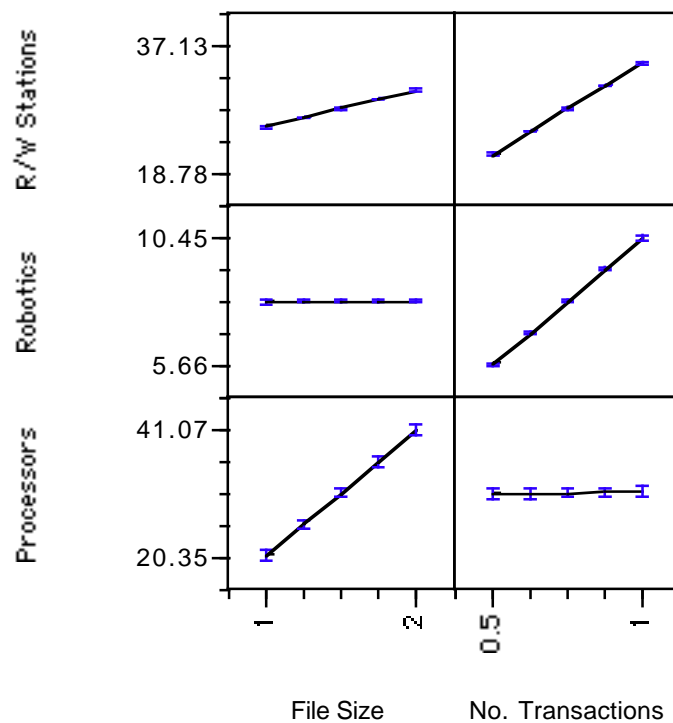
# Volume Composition results on a subset of resources



# Volume Composition network results

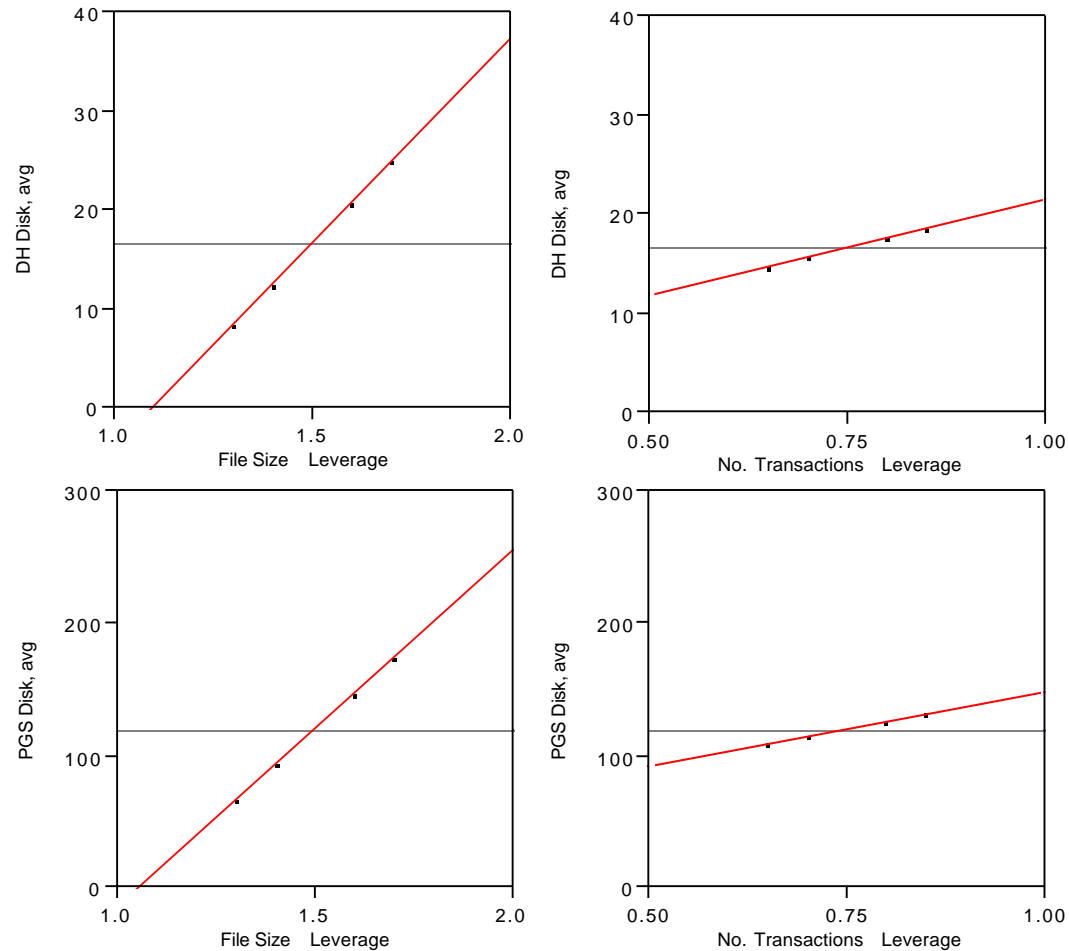


# Experiment Design allows statistical separation of variables



$$\text{Quantity} = \alpha * \text{File Size} + \beta * \# \text{ transactions}$$

# Effects on disk requirements confounded, but determinable



$$\text{Quantity} = \alpha * \text{Size} + \beta * \# \text{ transactions} + \gamma * \text{Size} * \# \text{ transactions}$$

# Impact of changing the Volume Composition



**Robotics needed changes almost 1-to-1 with number of requests**

- Robotics are transaction driven, not volume driven

**R/W Stations show ~70% change for a 100% change in number of requests**

- Indicates mixed transaction/volume dependencies, with transactions dominating
- R/W stations have a mount/search/rewind/dismount overhead that is significant compared to the time needed to read or write a single file

**Number of processors directly related to only the change in file size**

- Due to assumption that processing load is related to input file size(s)

# Impact of changing the Volume Composition (cont.)

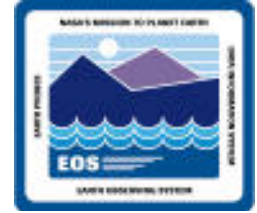


**Effects of number of transactions and file size on Disk requirements are confounded**

- In statistics “confounding” indicates an interaction between the variables
- Of the two variables, transaction size has the largest effect
- At DPS part of dependency on file size is attributable to the assumption relating processing load to input file size

**Network bandwidth demand scales directly with file size**

**Caveat:  $2^n$  experiments are used to determine which factors are important and any interaction between factors. To determine the functional relationship additional experiments would have to be performed.**



# Sensitivity to request type

## Experiment compares different mixes of user requests

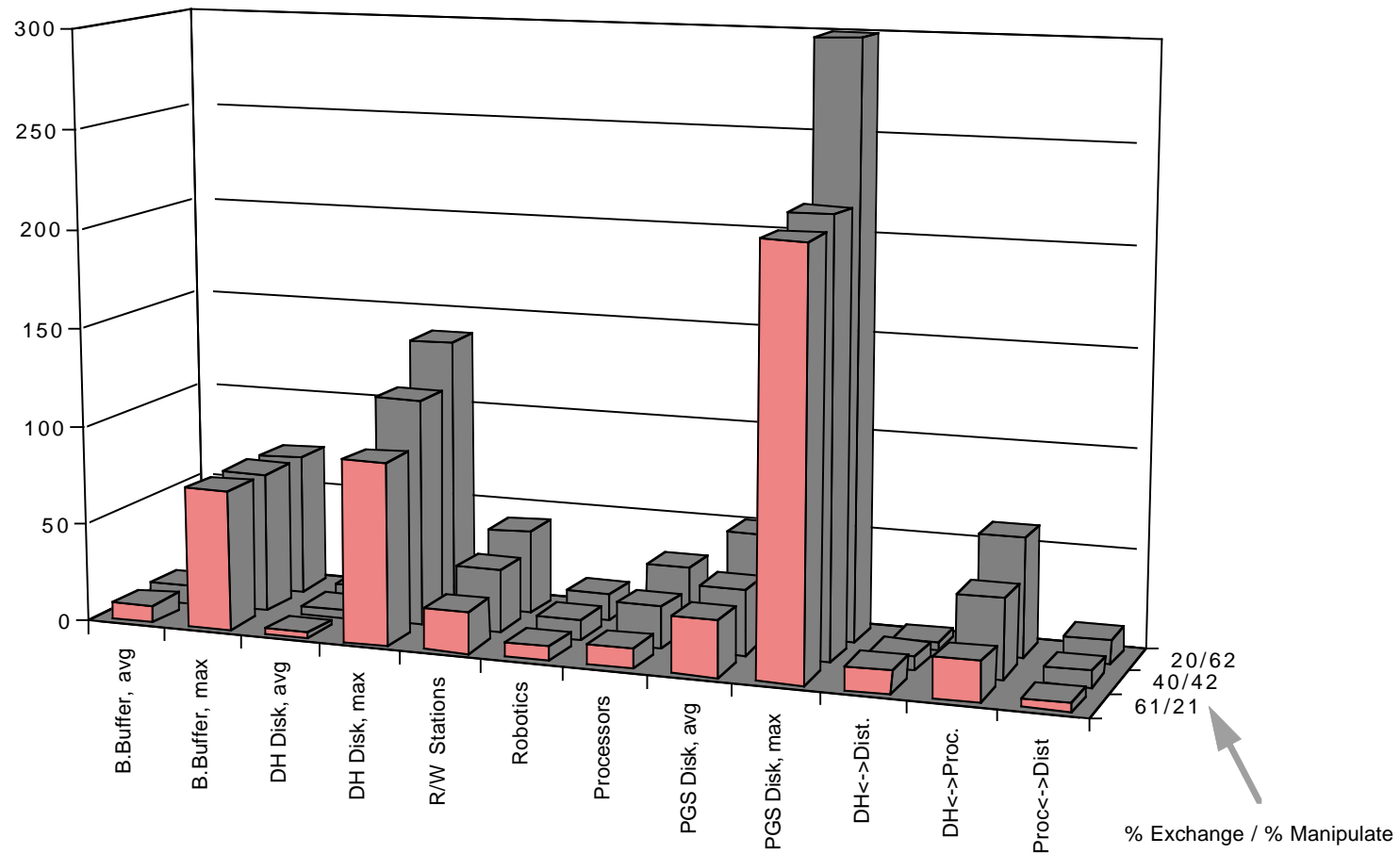
- **Exchange request extracts data from archive and transfers it to user**
  - subsetting, when present, based only on HDF structure
- **Manipulate request requires significant processing**
  - certain subsetting, on-demand processing, user supplied methods, ...
- **Three combinations of exchange and manipulate request types used**
  - 20%:62% , 40%:42% , 60%:22% (exchange:manipulate)
  - 40%:42% is baseline

## Assumptions:

- **Volume of data delivered to user and total number of requests held constant**
- **All other assumptions are as initially stated**



# Sensitivity to request type



# Results of sensitivity to request type



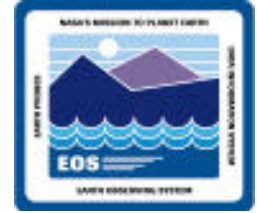
**Demand on all resources increase with number of Manipulate requests**

- **Robotics increases because on-demand and user-supplied methods are assumed to need multiple files per request**
- **R/W stations increases both because of multiple files per request and the decimation ratio**
- **Processor demand increase because of more manipulate requests**
- **Networks increase because of more manipulate requests**

**Sensitivity to request type is a candidate for further study**

- **Refinement of inputs from user model**

**Additional simulation based on analysis of uncertainty**



# Decimation Ratio Experiment

**Experiment compares differing decimation ratios and number of input files**

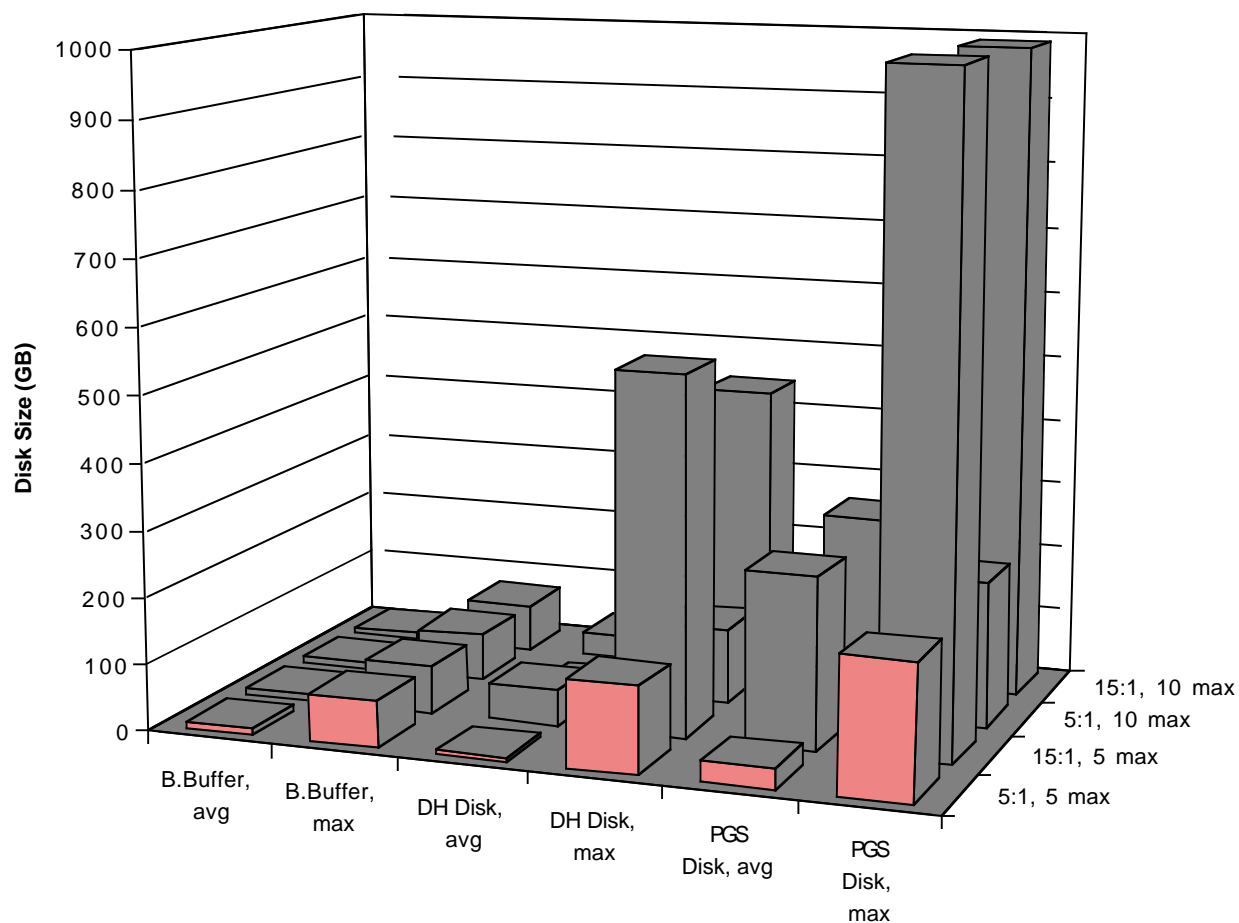
- Number of input files refers to the number of file used during a manipulate

**Experiment set up as a  $2^2$  full factorial statistically designed experiment**

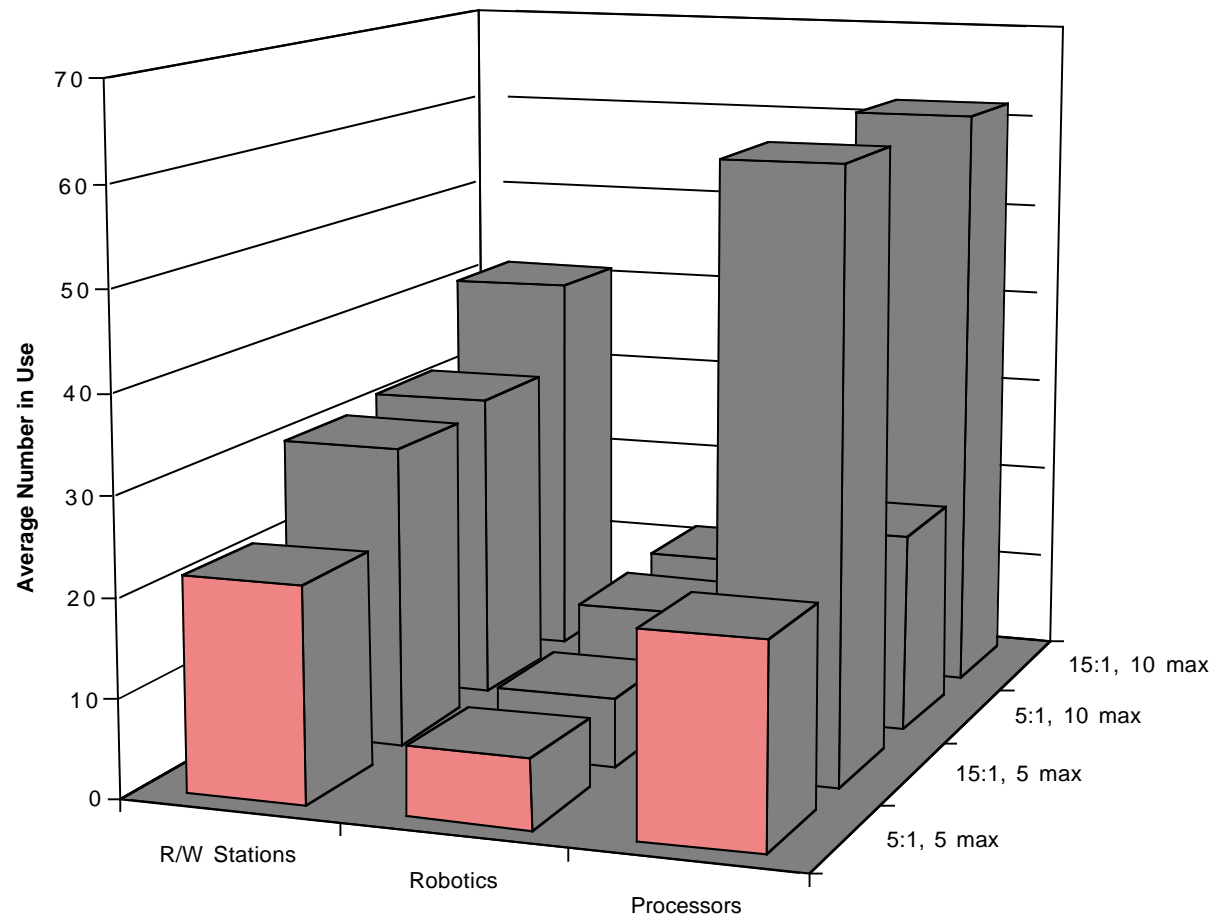
- Baseline decimation ratio (5:1) versus 3 times that (15:1)
- Baseline maximum number of input files (10) versus 1/2 that number (5)

**The decimation ratio is ratio of the volume of data extracted from the archive to the volume of data distributed to the users during manipulate requests.**

# Decimation disk requirement results

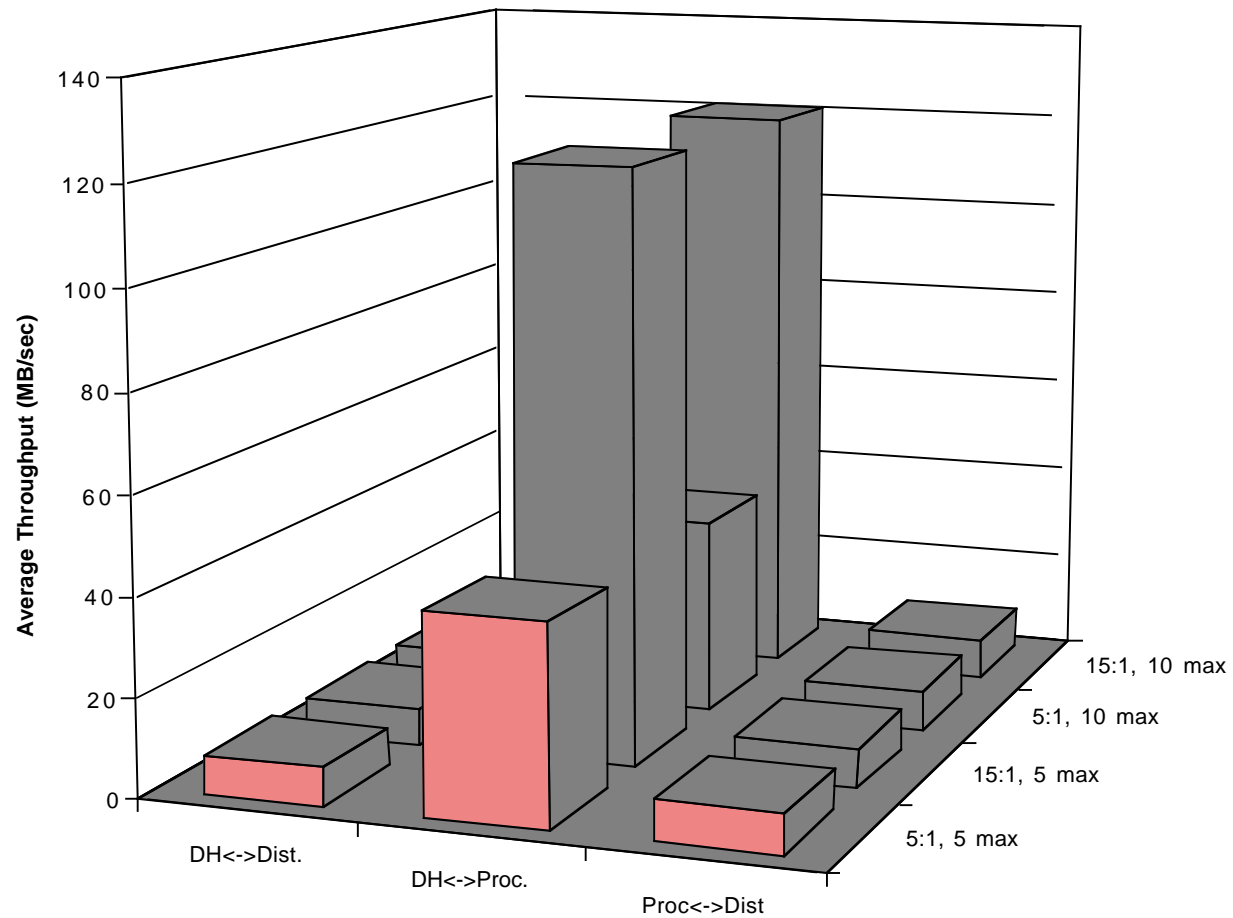


# Decimation results on a subset of resources

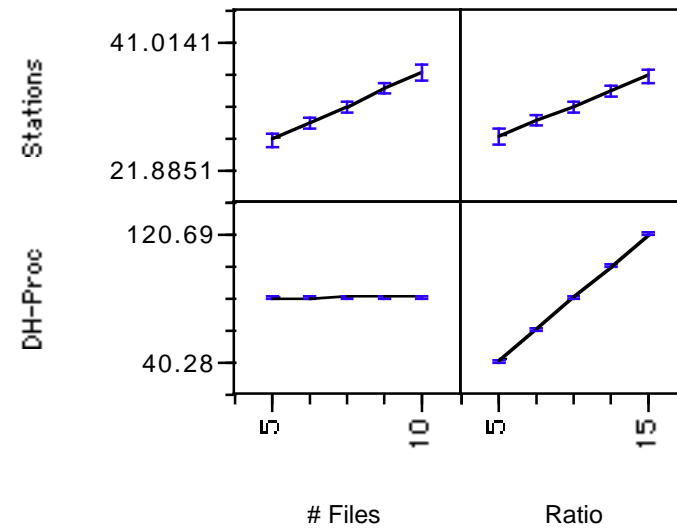
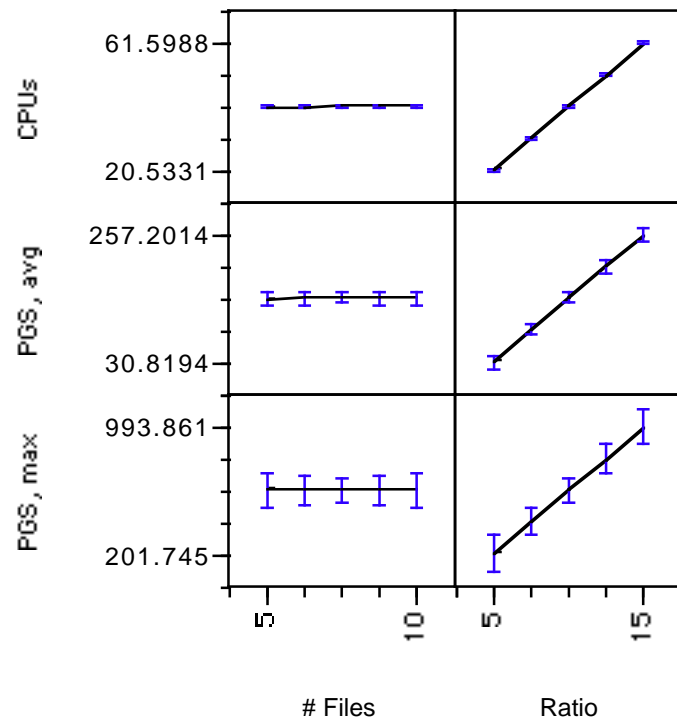




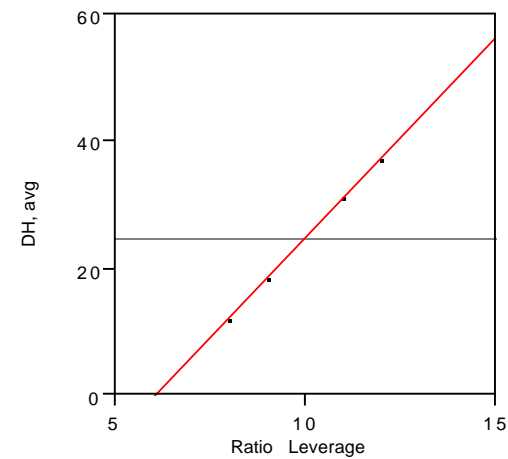
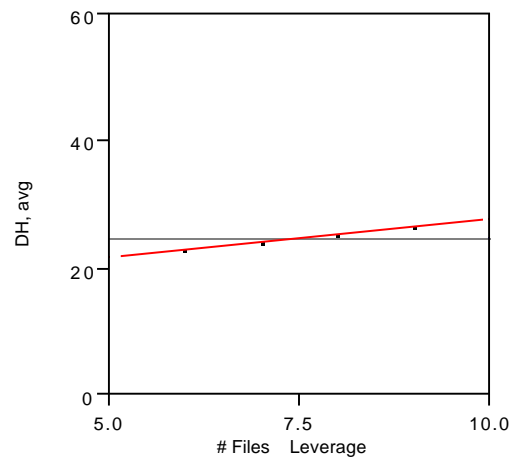
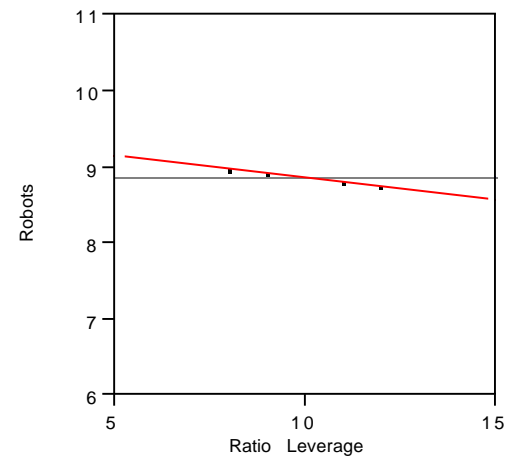
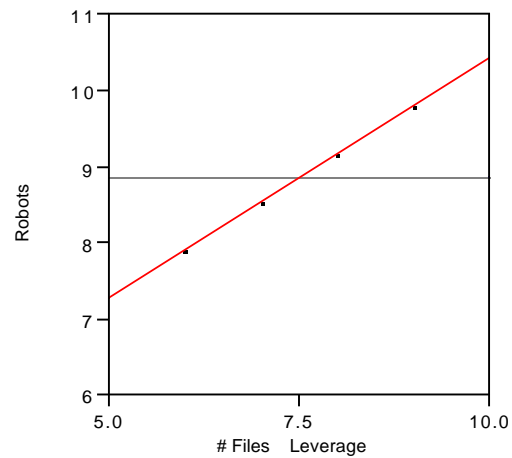
# Decimation network results



# Statistical separation of Decimation experiment variables



# Separation of confounded Decimation experiment variables





# Decimation Ratio Results

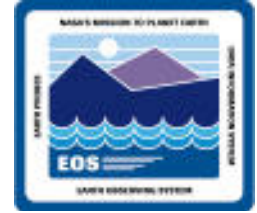
**Robotics depends on number of files, but not on decimation ratio**

- Manipulate request can involve retrieving multiple files
- Robotics are transaction driven, not volume driven

**R/W Stations show ~30% change for both decimation ratios and # of files**

**Working storage requirements largely dependent on decimation ratios**

- Effects both DSS and DPS
- Further experiment shows that the dependency is non-linear
- DSS shows a slight negative correlation with number of files
  - Probably related to network transfer times for large files



# Decimation Ratio Results (cont.)

## Number of processors related to only change in decimation ratio

- Due to assumption that processing load is related to input file size

## Network requirements mixed

- DSS $\leftrightarrow$ DPS directly related to decimation ratio
- All others are not

**Caveat:** 2<sup>n</sup> experiments are used to determine which factors are important and any interaction between factors. To determine the functional relationship additional experiments would have to be performed.

# Conclusions (to this point in the study)



**Archive robotics and r/w stations are sensitive to number of transactions**

- Assumptions relating to number of transactions should be refined/verified
- Suggests that techniques for lowering number of transactions into the archive should be investigated
  - Organizing data onto the data tapes and middle tier are possibilities
  - Must not offset gains due to blocking buffer

**Mix of types of user service requests should be refined and verified**

**Decimation ratio needs to be refined and verified**

**Assumption relating processing load to input size should be verified**

**Caveat: Additional concerns may be uncovered as further work is performed**